



De Faria Pires, Loïc. 2020. Influencia de la posesición de traducción automática en la calidad de los textos traducidos en la Dirección General de Traducción de la Comisión Europea. *Revista de lenguas para fines específicos* 26.1, pp. 26-38 · <http://dx.doi.org/XXXXXXXXXXXXX>
<https://ojsulpgc.es/ojs/index.php/LFE/index>

Paper received: 18 JAN 2019

Paper received in revised form and accepted for publication: 25 MAY 2019



Influencia de la posesición de traducción automática en la calidad de los textos traducidos en la Dirección General de Traducción de la Comisión Europea

Loïc De Faria Pires¹ ·  <https://orcid.org/0000-0002-9980-7175>

Université de Mons

Faculté de Traduction et d'Interprétation · Service de Traduction spécialisée et Terminologie (FTI-EII) · 17, Chemin du Champ de Mars · 7000 Mons · Belgique

RESUMEN

En el presente artículo, se presentan los resultados del primer proyecto piloto que se llevó a cabo en el marco de la investigación doctoral del autor. Más concretamente, se expondrán las conclusiones provisionales que se han sacado a partir del estudio de los 51 primeros textos de un corpus constituido exclusivamente de textos traducidos de inglés a francés procedentes de la Dirección General de Traducción (DGT) de la Comisión Europea. En dichos textos, se analizan los segmentos poseditados por los traductores institucionales voluntarios del proyecto en términos de modificaciones llevadas a cabo por los revisores de los departamentos de lengua francesa de la DGT: a cada modificación se le atribuye una penalización usando una herramienta de evaluación elaborada por Toudic et al. (2014), para calcular una penalización media para cada texto y para cada traductor. También se estudia, en dichos segmentos poseditados, la repartición de las modificaciones de los revisores en distintas categorías que describiremos brevemente. Finalmente, se expondrán unas conclusiones provisionales con respecto a la influencia de la posesición sobre la calidad de los textos de los distintos traductores estudiados.

Palabras claves: Posedición, Comisión Europea, DGT, Calidad, Corpus

ABSTRACT

The present article aims at presenting the results of the first pilot project we carried out in the framework of our Ph. D. research. In concrete terms, we will explain our provisional findings related to the study of the 51 first texts of a corpus constituted by texts translated from English into French in the European Commission's Directorate-General for Translation (DGT). In these texts, we analysed segments which were post-edited by the volunteer translators taking part in our project. We studied the modifications introduced in the post-edited segments by French units' revisers throughout the revision process. All modifications introduced by revisers in the post-edited segments were attributed a penalty score thanks to an evaluation tool issued by Toudic et al. (2014), so that we could compute an average penalty score for each text and translator. We also grouped all modifications by revisers in post-edited segments under several categories that we will briefly describe. Finally, we will draw provisional conclusions related to the influence of Machine Translation post-editing on the quality of the final texts provided by the translators participating in our project.

Keywords: Post-editing, European Commission, DGT, Quality, Corpus

1. Introducción

En el presente artículo se describen los resultados obtenidos tras el análisis de la primera parte del corpus de la investigación doctoral del autor. Dicha investigación se está llevando a cabo en un contexto institucional, y más

¹ Corresponding author · Email: Loic.DEFARIPIRES@umons.ac.be



concretamente en la Dirección General de Traducción (a continuación DGT) de la Comisión Europea. A pesar de la importancia fundamental de las actividades de la DGT en la aplicación de la legislación europea y en el acceso a los documentos europeos públicos de la Comisión por todos los ciudadanos europeos sin discriminación lingüística (Comisión Europea, 2014: 2), y del desarrollo rápido de la aplicación de posesición de traducción automática en dicho contexto de traducción institucional (Cadwell et al., 2016: 227), de momento se ha llevado a cabo muy poca investigación acerca de la calidad de los textos poseídos en la DGT. Hasta ahora, la mayoría de los estudios realizados en el contexto de la DGT se han centrado en el proceso (ver, p. ej, Cadwell et al., 2016) o en aspectos etnográficos de la profesión de traductor en esta institución (ver, p. ej, Koskinen, 2008). También se han redactado trabajos relativos a aspectos teóricos de la calidad de las traducciones en el marco de la DGT, por ejemplo en la obra muy completa *Quality aspects in Institutional Translation* editada por Svoboda et al. (2017). No obstante, no se pudo encontrar investigación relativa a un estudio concreto de textos traducidos y poseídos procedentes de la institución.

Este contexto institucional europeo constituye uno de los únicos en los que todavía no se ha llevado a cabo dicho tipo de investigación, en particular con respecto a la posesición de traducción automática. Pese a que numerosos estudios traductológicos realizados tienen como objeto de estudio el proceso de posesición (ver, p.ej, Carl et al., 2011; O'Brien, 2011; Koponen, 2012), también existen trabajos relativos al producto y a la calidad. De hecho, Martikainen y Kübler (2016) estudiaron la calidad de textos poseídos en el ámbito de la salud; Guerberof Arenas (2009), así como De Almeida y O'Brien (2010) analizaron la calidad de posesiciones de documentos relativos a las tecnologías de la información, en un contexto de localización; Peraldi (2016) llevó a cabo un estudio detallado sobre el tema de la posesición de textos financieros; y Koponen y Salmi (2015) así como Čulo et al. (2014) trabajaron sobre la calidad de la posesición de artículos de prensa, entre otros.

En aquel contexto de escasa investigación institucional sobre la calidad de la posesición, se decidió llevar a cabo una tesis doctoral relativa a la influencia de la posesición de traducción automática en la calidad de los textos producidos en la DGT. A continuación figuran los resultados del primer proyecto piloto realizado en el marco de dicha investigación. La hipótesis que se tomó como base para el presente trabajo está vinculada con los resultados obtenidos Čulo et al. (2014: 207): se pretende comprobar si, como concluyeron dichos autores, los traductores experimentados cometen tantos errores de posesición como los traductores principiantes. Para intentar lograr este objetivo, se analizan textos poseídos procedentes de la Dirección General de la Traducción de la Comisión Europea, y se intenta determinar los tipos de problemas de posesición que aparecen para varios participantes que se describirán a continuación.

2. Descripción del corpus

Cabe, en primer lugar, describir brevemente el corpus en el que se basa el presente artículo. Se trata de los primeros textos recopilados dentro del contexto de una investigación doctoral llevada a cabo por el autor. Más concretamente, dicho corpus provisional está compuesto por 51 textos procedentes de la DGT, repartidos entre 6 traductores del departamento de lengua francesa (cinco funcionarios experimentados y un traductor *Blue Book*², considerado en el presente estudio como poco experimentado). Todos los textos usados se poseitaron

² “The Blue Book Traineeship scheme offers a paid internship of 5 months with the European Commission or some executive bodies of the European Institutions like the European External Action Service, starting on either 1st March or 1st October. [...] The traineeship program is open to university graduates from all over the world who have a:

de inglés a francés. De hecho, para la tesis doctoral en general y el presente artículo en particular, sólo se estudian los segmentos poseditados, es decir los segmentos de traducción automática procedentes del motor estadístico *Moses* usado en este departamento lingüístico, que poseditaron los traductores voluntarios de la DGT que participan en el proyecto. Para aclarar la noción de posesición, una definición posible sería la de Sharon O'Brien: "*Post-editing is the correction of raw machine translated output by a human translator according to specific guidelines and quality criteria*" (2011: 197). Por lo que se refiere a los tipos de textos, existe mucha variación en el corpus de estudio: se tratan tanto de textos legislativos como de documentos internos, comunicaciones para el público, documentos de trabajo, discursos de las autoridades...

A continuación se presentan la metodología y las herramientas a las que se recurre en el marco del presente artículo.

3. Metodología y herramientas

3.1. Herramienta de evaluación

En el marco de la presente investigación, se recurre a una plantilla de evaluación de la calidad en traducción elaborada por los investigadores franceses Toudic et al. (2014). Esta plantilla permite analizar con precisión la calidad de las partes poseditadas de cada texto, y permite definir categorías de modificaciones, que se describirán más adelante. Dicha plantilla define 6 niveles de gravedad para cada modificación (del 0 [efecto imperceptible] al 5 [efecto muy grave]), así como 9 categorías de modificaciones (fidelidad, adición y omisión de información, terminología, fraseología, gramática y sintaxis, ortografía y tipografía, estilo, localización y publicación asistida por ordenador).

Antes de analizar los resultados obtenidos a partir de dicha plantilla, era necesario comprobar que la herramienta proveía resultados fiables, así que se calculó un coeficiente de correlación para las dos medidas llevadas a cabo, las cuales se describirán a continuación.

3.2. Modificaciones por los revisores y penalizaciones

El primer tipo de medidas que se realizó para el presente artículo fue, como en el marco de la tesis, una medida de penalización destinada a evaluar la calidad de los segmentos poseditados en cada texto. Cabe destacar, en primer lugar, que se calculó un coeficiente de tipo kappa de Cohen ponderado lineal para comprobar que la atribución de las penalizaciones fuera fiable. Así pues, el autor llevó a cabo este cálculo con otra evaluadora con el mismo grado de experiencia, y se obtuvo un valor de 0,8919 para la atribución de las penalizaciones, lo que corresponde a una concordancia excelente entre ambos evaluadores (Santos, 2017: 2). Significa que, en la mayoría de los casos, cada evaluador le atribuyó la misma penalización a cada modificación, o penalizaciones muy parecidas.

1. degree of at least 3 years of study (minimum a bachelor),

2. very good knowledge of English or French or German.

3. very good knowledge of a second EU official language (required for nationals of EU countries)." (European External Action Service, 2016)

Se describe a continuación el procedimiento exacto: en primer lugar, se aislaron los segmentos poseditados en los 51 textos que la DGT ya le había transmitido al autor. En todos estos segmentos, se recopilaron las modificaciones que introdujeron los revisores (en la DGT se revisa cada texto tras su traducción / posesición) en las versiones poseditadas por los traductores voluntarios. Una vez listadas todas las modificaciones de los revisores en los segmentos poseditados, a cada modificación se le atribuyó una penalización tomando como base la plantilla de evaluación de Toudic et al. (2014), que se describe arriba. De este modo, se obtuvo una penalización global para cada texto. No obstante, se tiene que tener en cuenta el hecho de que todos los textos constitutivos del corpus provisional llevan un número de palabras distinto. Fue por ello por lo que se decidió calcular, para cada texto, una penalización normalizada por franja de 100 palabras en los segmentos poseditados. De este modo, fue posible comparar la calidad de la posesición para cada texto analizado.

Tras obtener una puntuación global para cada texto investigado de manera individual, se calculó una penalización media para cada traductor, con el fin de poder comparar la calidad de las posesiciones llevadas a cabo por distintos traductores en este contexto institucional europeo, que es el objeto de estudio principal de la investigación doctoral descrita en el marco del presente artículo.

Parece fundamental destacar que, a pesar de que en el ámbito de la traductología se hable de “errores”, en este proyecto se habla de “modificaciones”, porque todas las modificaciones introducidas por los revisores en los segmentos poseditados no necesariamente son errores, sino también modificaciones preferenciales (relativas a la formulación, por ejemplo).

A continuación se detallará la siguiente etapa del análisis de nuestro corpus provisional, es decir la clasificación en distintas categorías de las modificaciones introducidas por los revisores en los segmentos poseditados.

3.3. Categorías de modificaciones

Tras haber contabilizado todas las modificaciones por los revisores y calculado una penalización media para cada texto y cada traductor, también se decidió clasificar todas las modificaciones introducidas por los revisores en los segmentos poseditados en distintas categorías, basándose en las categorías propuestas por Toudic et al. (2014). Para verificar la fiabilidad de esta etapa, se calculó un valor de correlación de tipo kappa de Cohen clásico (es decir binar) con la investigadora a la que se presentó precedentemente, y se obtuvo un valor de acuerdo de 0,7377, lo que corresponde a un acuerdo satisfactorio (Santos, 2017: 2), que es muy aceptable si tenemos en cuenta las 9 categorías posibles para cada modificación.

De este modo, resultó posible definir una repartición de las modificaciones de los revisores en dichas distintas categorías para cada texto y cada traductor: se calculó el número de modificaciones de cada tipo por franja de 100 palabras por texto y por traductor, con el objetivo de poder comparar los textos poseditados entre sí, y se procedió del mismo modo con las penalizaciones medias atribuidas a cada categoría para cada texto y cada traductor.

4. Resultados provisionales

A continuación, se exponen y se interpretan los resultados provisionales. Se reparten en tres secciones: las penalizaciones, las reparticiones de las modificaciones por los revisores en categorías y las reparticiones de las penalizaciones en categorías.

4.1. Penalizaciones obtenidas

En primer lugar se analizan y se comparan las penalizaciones medias obtenidas por cada traductor por franja de 100 palabras, con el objetivo de determinar si aparecen diferencias notables de calidad de posesición entre los distintos traductores. Los resultados presentados en este artículo todavía no serán significativos, puesto que sólo se dispone de datos para 5 traductores experimentados y 1 traductor poco experimentado, lo que no es suficiente para poder comparar las tendencias de grupo. Sin embargo, ya se observan muchas variaciones individuales, como aparece en la siguiente tabla.

Traductor	Penalización media / 100 palabras poseídas
Experimentado 1	1,7574
Experimentado 2	1,4973
Experimentado 3	0,7468
Experimentado 4	3,3456
Experimentado 5	1,6087
Poco experimentado 1 (<i>Blue Book</i>)	3,3616

Como se muestra en la tabla, todos los participantes tienen ciertos problemas de posesición, puesto que todos obtienen una penalización (una posesición sin modificaciones por los revisores obtendría una penalización nula). Sin embargo, de momento no se puede concluir que todos los traductores institucionales investigados cometan tantos “errores” de posesición como dicen Čulo et al. (2014) (lo que también verificaremos en el punto 4.2), o por lo menos que la posesición tenga el mismo impacto en la calidad del trabajo de los traductores participantes.

De hecho, si es verdad que el traductor poco experimentado obtiene la penalización más alta, uno de los traductores experimentados (el número 4), obtiene una penalización similar. No obstante, aparece mucha variación entre los traductores, sean poco experimentados o no. El traductor experimentado 3 obtiene una penalización media por franja de 100 palabras de 0,7468 en sus segmentos poseídos, es decir casi cinco veces inferior a la penalización obtenida por el traductor experimentado 4 y el traductor poco experimentado, lo que representa una diferencia considerable.

Entre estos extremos, aparecen penalizaciones medias que oscilan entre 1,4973 y 1,7574, es decir bastante similares, así que de momento se puede deducir de manera muy provisional que la posesición tiene efectos similares en la calidad de los textos proveídos por los traductores experimentados 1, 2 y 5.

En la sección siguiente se analiza el número de modificaciones introducidas por los revisores en las posesiciones de cada traductor, y se intenta establecer una correlación con las penalizaciones descritas arriba.

4.2. Número de modificaciones

Después de haber calculado las penalizaciones medias para los segmentos poseídos de cada texto y cada traductor, también se calculó el número medio de modificaciones introducidas por los revisores, con el objetivo de poder comparar estas dos series de datos y determinar las relaciones que las unen. Así pues, se compuso la misma tabla que arriba, sustituyendo las penalizaciones medias por los números medios de modificaciones.

Traductor	Número medio de modificaciones / 100 palabras poseditadas
Experimentado 1	1,1394
Experimentado 2	1,0159
Experimentado 3	0,4994
Experimentado 4	2,1632
Experimentado 5	1,0355
Poco experimentado 1 (<i>Blue Book</i>)	1,9006

Como se puede ver, la repartición sigue bastante parecida, excepto por lo que se refiere al traductor experimentado 4 y al traductor poco experimentado, fenómeno para el que se intentará formular una explicación a continuación.

En principio, cabe destacar que también se observa en el caso del número medio de modificaciones el mismo fenómeno de variación individual. De hecho, mientras que los textos poseditados procedentes del traductor experimentado 3 sólo lleven una media de 0,4994 modificación por franja de 100 palabras, los del traductor experimentado 4 llevan un número medio de modificaciones más de cuatro veces superior. Los textos poseditados por el traductor poco experimentado tienen una media de 1,9006, es decir un número bastante elevado, sin ser el extremo alto. Por su parte, los traductores experimentados 1, 2 y 5 obtienen resultados similares, fenómeno similar al que ya se producía con las penalizaciones medias.

Cabe volver a hablar del cambio que se observa para el traductor experimentado 4 y el traductor poco experimentado entre las penalizaciones medias por franja de 100 palabras (medida para la que el traductor experimentado 4 obtiene 3,3456 y el traductor poco experimentado 3,3616) y el número medio de modificaciones por 100 palabras poseditadas (2,1632 para el traductor experimentado 4 y 1,9006 para el traductor poco experimentado). Esta inversión entre los valores permite poner de relieve otro factor importante en el marco del presente trabajo: la gravedad de los elementos causadores de las modificaciones por los revisores.

A primera vista, el hecho de que el traductor experimentado 4 obtenga una penalización media inferior a la del traductor poco experimentado a pesar de obtener un número medio de modificaciones superior puede extrañar. Sin embargo, la explicación está sin duda vinculada a la gravedad de las modificaciones: a pesar de que aparezcan más modificaciones en los textos poseditados por el traductor experimentado 4, pueden ser menos graves, lo que justificaría su penalización media inferior. En la siguiente sección, se tratará de confirmar esta hipótesis comparando la repartición de las penalizaciones y del número de modificaciones en las distintas categorías de la plantilla de Toudic et al. (2014).

4.3. Repartición de las modificaciones y de las penalizaciones entre las distintas categorías

Para intentar determinar el impacto de las distintas categorías de modificaciones en la calidad final de los textos poseditados en la DGT, es necesario estudiar la proporción de cada tipo modificación en el número total de modificaciones por los revisores y en las penalizaciones medias de los textos poseditados por cada traductor. A continuación aparece una tabla que tiene como objetivo comparar estos elementos: para cada traductor, se calculó la repartición de las categorías de modificaciones con respecto a las penalizaciones (P) y al número de modificaciones (M) en los segmentos poseditados.

Fidelidad		Adición/Omisión		Terminología		Fraseología		Gramática/Sintaxis		Ortografía/Tipografía		Estilo		Localización		PAO		
P	M	P	M	P	M	P	M	P	M	P	M	P	M	P	M	P	M	
Exp1	<u>36%</u>	26%	<u>9%</u>	7%	21%	<u>22%</u>	21%	<u>26%</u>	3%	<u>5%</u>	3%	<u>5%</u>	6%	<u>9%</u>	0%	0%	1%	1%
Exp2	<u>31%</u>	21%	<u>10%</u>	8%	18%	<u>20%</u>	19%	<u>23%</u>	4%	<u>5%</u>	10%	<u>12%</u>	8%	<u>11%</u>	0%	0%	0%	0%
Exp3	<u>34%</u>	20%	<u>2%</u>	1%	<u>23%</u>	18%	14%	<u>21%</u>	4%	<u>6%</u>	14%	<u>21%</u>	2%	<u>3%</u>	7%	<u>10%</u>	0%	0%
Exp4	<u>41%</u>	27%	7%	<u>8%</u>	23%	<u>27%</u>	20%	<u>26%</u>	0%	0%	5%	<u>7%</u>	4%	<u>5%</u>	0%	0%	0%	0%
Exp5	<u>27%</u>	14%	<u>14%</u>	12%	8%	<u>10%</u>	32%	<u>36%</u>	6%	<u>8%</u>	8%	<u>12%</u>	4%	<u>6%</u>	0%	0%	1%	<u>2%</u>
P.ex	<u>46%</u>	29%	<u>8%</u>	7%	14%	<u>16%</u>	16%	<u>24%</u>	4%	<u>6%</u>	6%	<u>11%</u>	4%	<u>6%</u>	<u>2%</u>	1%	0%	0%

Por ejemplo, si se consideran todos los segmentos poseídos por el traductor experimentado 1 (Exp 1) y que se analiza el componente de fidelidad, se observa que las penalizaciones relativas a la fidelidad constituyen un 36% de la penalización global obtenida por este traductor, mientras que, en términos de número de ocurrencias, las modificaciones relativas a este componente representan un 26% del total de modificaciones por los revisores para este traductor. El valor superior de las penalizaciones con respecto al número de modificaciones significa que se les atribuyó más penalizaciones a las modificaciones relativas a la fidelidad que, por ejemplo, a las modificaciones relativas a la terminología para este mismo traductor (21% de la penalización total media, y 22% del número medio de modificaciones). Al analizar la diferencia entre ambos datos, se puede deducir la gravedad relativa de las modificaciones pertenecientes a cada categoría.

En cada par de columnas, los elementos subrayados permiten indicar la proporción más elevada para cada categoría en términos de penalizaciones o de número de modificaciones, lo que permite estimar la gravedad relativa de cada tipo de modificación. Cuando el elemento subrayado se encuentra en la columna de la izquierda para un tipo dado de modificación, significa que la gravedad relativa de las modificaciones de este tipo es más elevada que la de otros tipos de modificaciones para las que la proporción en términos de número de modificaciones es más alta que su proporción en términos de penalización.

Los resultados expuestos en la tabla permiten suponer que las modificaciones por los revisores en los segmentos poseídos relativas a la fidelidad y a la adición/omisión de informaciones (número de modificaciones < penalización) tienen un efecto más grave en la calidad final de las posesiciones institucionales estudiadas que las modificaciones relativas a las demás categorías (número de modificaciones > penalización), pero habrá que confirmar esta suposición y verificarla de manera más sistemática en el marco del proyecto, cuando se obtengan más datos.

Acerca de la inversión entre penalizaciones y número de modificaciones entre el traductor experimentado 4 y el traductor poco experimentado, a partir de lo que se dedujo, y a partir de lo que se puede observar en la tabla, se nota que el traductor poco experimentado tuvo más problemas relativos a la fidelidad que el traductor experimentado 4. En cambio, se observa que, aunque las posesiciones del traductor experimentado 4 hayan sufrido más modificaciones en términos de número, se tratan de modificaciones que pertenecen a categorías que

tienen menos impacto en la calidad final de los textos poseditados (p. ej. más modificaciones relativas a la terminología).

Cabe destacar que la definición y delimitación de las nueve categorías de modificaciones figuran en el artículo de Toudic et al. (2014: 9), por lo que no se explicarán en detalle las consideraciones teóricas relativas a dichas modificaciones en el marco del presente artículo, y también cabe destacar que la repartición de las modificaciones entre las distintas categorías permiten establecer un orden entre dichas categorías, para observar los tipos de modificaciones que más se producen, como se puede ver en la tabla siguiente.

Categoría	Intervalo de proporciones (en términos de número de modificaciones)
1) Fraseología	21-36 %
2) Fidelidad	14-27 %
3) Terminología	10-27 %
4) Ortografía / Tipografía	5-21 %
5) Estilo	3-11 %
6) Adición / Omisión	1-12 %
7) Localización	0-10 %
8) Gramática / Sintaxis	0-8 %
9) PAO	0-2 %

Por supuesto, las reparticiones de los números de modificaciones varían entre los distintos participantes (por ejemplo, la proporción de modificaciones relativas a la terminología más baja es de 10 % para el traductor experimentado 5, y la más elevada es de 27 % para el traductor experimentado 4), pero se pueden extraer tendencias, que se tratará de sistematizar conforme irá avanzando la tesis doctoral descrita en el presente artículo. De hecho, sería interesante comparar las reparticiones de las modificaciones por los revisores en función de los distintos traductores y de los distintos tipos de documentos traducidos y/o poseditados en la DGT, basándose por ejemplo en la tipología relativa a las exigencias de calidad elaborada por la DGT (Dirección General de Traducción, 2015).

Finalmente, cabe destacar que los resultados e interpretaciones descritos arriba son muy provisionales, y que se seguirán recopilando datos durante los meses que vienen (por lo menos hasta finales del 2019). Además, está planeado llevar a cabo una experimentación de posesición en nuestra facultad con estudiantes de último año de Máster, usando los textos poseditados por los traductores institucionales voluntarios, con el fin de obtener datos de una categoría de traductores que todavía no hemos investigado, es decir los estudiantes en traducción.

5. Limitaciones

Por supuesto, los resultados presentados en este artículo todavía conllevan unas limitaciones, puesto que lo que se presenta aquí consiste en un resumen del primer proyecto piloto que se llevó a cabo con el fin de validar la metodología provisional del proyecto. Se tratará, en la medida de lo posible, de solucionar estos límites a corto y medio plazo, pero de momento se deben tener en cuenta, porque son inherentes a los datos que se han obtenido hasta ahora y al contexto en el que se está llevando a cabo el proyecto.

5.1. Datos insuficientes

La relativa pequeña cantidad de datos que se ha recopilado hasta ahora constituye la mayor limitación actual del proyecto. Se empezó a recoger textos a finales del año 2017 y, en un plazo de un año, ha sido posible obtener 51 textos.

Así pues, puesto que el autor pretende seguir recopilando textos durante por lo menos un año, se puede imaginar que se podrán obtener por lo menos otros 50 textos procedentes de los mismos traductores experimentados y de otros traductores *Blue Book*.

Por supuesto, resulta muy difícil determinar con precisión la cantidad total de textos que se logrará obtener: dado el contexto institucional de la DGT en el que se trabaja, contexto en el que los traductores están sometidos a un nivel de presión laboral bastante elevado (Comisión de las Comunidades Europeas, 2006: 4), les resulta difícil mandar textos nuevos con regularidad, así que la cantidad final de datos está, en cierta medida, pendiente de su trabajo y de su voluntad de tomar el tiempo necesario para mandarle al autor documentos que analizar. Como consecuencia, no se dispone de un control total sobre la cantidad de textos que compondrán el corpus final, a pesar de que se trate de mantener un contacto constante con los voluntarios de la Comisión Europea que se encargan de la organización logística del proyecto.

Se intentará pues componer un corpus tan voluminoso como posible, sin olvidar que en este tipo de contextos institucionales resulta imposible constituir un corpus que represente la totalidad de los traductores y de la producción de textos, lo que no necesariamente constituye un obstáculo según Koskinen: *“By focusing on a detailed analysis of small samples it is, I trust, possible to increase our understanding of the larger whole the samples were abstracted from. And by keeping the sample sizes reasonable, it is possible to accomodate several kinds of data and methods”* (2008: 7).

5.2. Cantidad de textos por traductor

Otra limitación del presente artículo consiste en la falta de equivalencia entre las cantidades de textos de los que se dispone para cada traductor. Por supuesto, ya se han estudiado varios textos poseditados por traductor, y desde luego de datos suficientes para sacar las primeras conclusiones que se expusieron en el presente artículo, pero está claro que para constituir un corpus doctoral relevante es necesario disponer de cantidades de datos similares para cada traductor, con el fin de poder llevar a cabo comparaciones relevantes.

Así pues, una de las misiones principales durante los próximos meses consistirá en equilibrar la cantidad de datos entre los distintos traductores que participan al proyecto, con el fin de poder validar los resultados de manera científica.

5.3. Tipos de documentos distintos para cada traductor

La tercera limitación que hay que tomar en cuenta es la no correspondencia entre los tipos de documentos poseditados por los traductores investigados. Aparece necesario destacar que, en el ámbito institucional de la DGT, se traducen numerosos tipos de documentos: legislación, comunicaciones, documentos internos de trabajo... y que cada traductor traduce principalmente documentos relativos a un asunto institucional

correspondiente a área de especialización, como los asuntos exteriores o la movilidad, por ejemplo (Comisión Europea, 2014: 4).

Si se considera lo que precede y se añade el carácter voluntario de la participación de los traductores al proyecto, fácilmente se puede imaginar que resulta totalmente imposible disponer de textos de ámbito equivalente para todos los traductores: se trata de una limitación inevitable inherente al contexto auténtico de nuestro proyecto, por lo que de momento se tendrá que aceptarla en calidad de variable de confusión (Shuttleworth, 2018). Sin embargo, todos los textos tienen el punto común de ser textos institucionales. Además, a cada documento se le otorga una letra correspondiente a su tipología (A, B, C o D) que determina el proceso de revisión que seguir (Dirección General de Traducción, 2015: 4), así que se podrán aplicar niveles de tolerancia distintos en función del tipo de texto, para poder comparar los textos poseditados teniendo en cuenta esta distinción. También se tratará de una de las etapas posteriores del proyecto, puesto que se intentará determinar si los distintos niveles de revisión en función de los tipos de documentos tienen un efecto sobre la calidad final de los segmentos poseditados. De este modo, la no correspondencia entre la tipología de los textos proveídos por cada traductor ya no constituirá una variable de confusión, sino una variable independiente.

5.4. Necesidad de verificar estas primeras tendencias

También parece imprescindible destacar que todo lo que se ha expuesto en presente artículo no son más que tendencias que fue posible interpretar gracias al análisis del corpus de investigación provisional. Así pues, se tendrá que llevar a cabo un análisis científico más completo y exhaustivo cuando se obtenga un número de textos suficiente como para poder sacar conclusiones relevantes, pero sí parecía interesante presentar aquí los primeros resultados de la investigación, así como el procedimiento metodológico general.

5.5. Falta de comparación con la traducción automática en bruto

A pesar de que el presente proyecto tenga como objetivo principal evaluar la calidad global de los segmentos poseditados por distintos traductores y grupos de traductores tras un análisis de las revisiones, también parece interesante afinar dicha metodología e intentar determinar cuáles son las modificaciones por los revisores que provienen directamente de la traducción automática en bruto propuesta por el motor de traducción de la DGT. De hecho, los revisores a veces modifican los elementos introducidos por el traductor durante el proceso de posesición, es decir elementos que no provienen directamente de la traducción automática en bruto pero sí ocurren durante el proceso de posesición (que es el que se estudia en el marco de nuestra investigación). Sería interesante diferenciar las marcas de revisión en las posesiciones que provienen directamente de la traducción automática en bruto y las que no proceden de ella.

5.6. Distintos revisores para cada texto

En el marco de nuestro proyecto, será necesario lidiar con otra variable de confusión: aparecen muchos revisores distintos en nuestro corpus. Aunque trabajemos con una cantidad significativa de textos procedentes de los mismos traductores, un elemento sobre el que no se tiene ninguna influencia son los revisores que se encargan de la revisión. También se podría estimar de manera aproximativa la severidad de cada revisor en respecto a los textos que revisa, pero hacerlo no entra en el marco de la presente investigación.

No obstante, se ha de encontrar una manera de sortear este problema, y quizás la manera más lógica de hacerlo sería suponer de manera teórica que cada texto traducido y/o poseditado en la DGT debe alcanzar un nivel de calidad determinado para su publicación y respetar los mismos criterios de calidad (ver, p. ej., Dirección General de Traducción, 2015) y que los revisores, independientemente de su identidad individual, llevan todos a cabo la misma misión: preparar los textos traducidos y/o poseditados para su futura publicación en el marco de las actividades de la Comisión Europea.

6. Conclusiones

Cabe ahora concluir brevemente la presente contribución. En el marco de este artículo, se introdujo el contexto de la investigación doctoral del autor. Después, se presentó el corpus provisional de investigación, así como las herramientas que se usaron en el marco del análisis de este corpus. También se describió la metodología, es decir el análisis de las revisiones en los segmentos poseditados y la evaluación de su calidad gracias al número de modificaciones, a la atribución de penalizaciones y a la clasificación de las modificaciones en distintas categorías, todo aquello usando la plantilla de evaluación de Toudic et al. (2014).

Con respecto a los resultados provisionales obtenidos, en primer lugar se pudo deducir que de momento no necesariamente aparece la equivalencia descrita por Čulo et al. (2014: 207), que decían que los traductores poco experimentados y los traductores experimentados cometían tantos errores de posesición (o, en la terminología del presente proyecto, recibían tantas “modificaciones”), puesto que tanto en términos de número de errores como en términos de penalización aparecen variaciones individuales bastante marcadas, que no parecen depender totalmente del grado de experiencia. De hecho, el traductor poco experimentado obtiene aproximadamente las mismas ponderaciones que el traductor experimentado 4, mientras que el traductor experimentado 3 obtiene valores mucho más bajos que los demás, y que los traductores 1, 2 y 5 producen resultados bastante homogéneos, tanto en términos de número de modificaciones como en términos de penalizaciones globales.

Al comparar las diferencias entre las medias relativas al número de modificaciones para cada traductor y las penalizaciones atribuidas a estas modificaciones en función de las categorías a las que pertenecen dichas modificaciones, se pudo deducir la gravedad relativa de cada categoría de modificaciones. De este modo, fue posible destacar que en el marco de los textos institucionales poseditados analizados generalmente parecen más graves las modificaciones por los revisores relativas a la fidelidad y a la omisión de informaciones (proporción de número de modificaciones inferior a la proporción de penalizaciones) que las modificaciones pertenecientes a las demás categorías como la terminología, la fraseología y la sintaxis, por ejemplo (proporción de número de modificaciones superior a la proporción de penalizaciones).

Además, se pudo observar que aparece un orden bastante delimitado con respecto a la distribución de las modificaciones por los revisores en los segmentos poseditados en las distintas categorías por los autores de la herramienta de evaluación usada en el marco del proyecto.

Finalmente, cabe hacer hincapié en el hecho de que subsisten muchas limitaciones en el marco de este artículo, que se tratará de solucionar de la manera más adecuada posible conforme vaya avanzando la investigación doctoral descrita en el artículo. Cada investigación empírica supone limitaciones inherentes a la metodología empleada, al contexto de estudio, y a numerosos otros factores. Más particularmente, el contexto institucional en el que se trabaja aquí constituye un entorno particular y bastante aislado, distinto del mundo profesional

general y más difícil de acceder. Así pues, en un contexto de este tipo, le resulta más complicado al investigador obtener textos y datos, llevar a cabo encuestas, obtener ciertos documentos de referencia y mantener una comunicación constante. No obstante, también le resulta muy retador e interesante al investigador poder llevar a cabo su investigación en un contexto tan desconocido, trabajando con un recurso tan novedoso como la traducción automática y en ámbito casi nunca explorado en el contexto de las instituciones internacionales: la posesición.

Así pues, las etapas posteriores de la investigación consistirán principalmente en seguir con la constitución del corpus de estudio, puesto que el objetivo es llegar a componer un corpus explotable antes de finales del 2019. Entonces, se tratará de obtener un número similar de textos para cada traductor experimentado voluntario, de recopilar textos procedentes de otros traductores poco experimentados de tipo *Blue Book*, y también de recopilar textos institucionales poseditados procedentes de estudiantes de Master de nuestra facultad, la Facultad de Traducción e Interpretación de la Universidad de Mons. Además, se intentará verificar las conclusiones provisionales expuestas en el presente trabajo, y llevar a cabo comparaciones entre la traducción automática en bruto, las posesiciones y las revisiones de cada texto poseditado, con el fin de poder determinar las modificaciones que directamente resultan del uso de la traducción automática en bruto.

About the author

Loïc de Faria Pires obtuvo un Máster de Traducción Especializada en el 2016 (EN-ES-PT > FR) y un Máster de Didáctica de las lenguas en el 2017, en la Universidad de Mons (Bélgica). Está llevando a cabo una investigación doctoral en el departamento de Traducción especializada y Terminología de la Facultad de Traducción e Interpretación de la misma universidad. Dicha investigación doctoral tiene como objeto de estudio la calidad de los contenidos poseditados de inglés a francés en el departamento de lengua francesa de la Dirección General de Traducción de la Comisión Europea en función del nivel de experiencia profesional de los traductores. Aunque su investigación doctoral esté centrada en el producto de posesición, también ha llevado a cabo trabajos relativos al proceso de posesición. Sus intereses principales son las tecnologías de traducción, la traducción automática y su posesición, y la traducción institucional.

Bibliografía

- Cadwell, P., Castilho, S., O'Brien, S., & Mitchell, L. (2016). Human factors in machine translation and post-editing among institutional translators. *Translation spaces*, 5(2), 222-243.
- Carl, M., Dragsted, B., Elming, J., Hardt, D., & Jakobsen, A. L. (2001). The process of post-editing: a pilot study. *Proceedings of the 8th International NLPSC workshop*, 41, 131-142.
- Comisión de las Comunidades Europeas. (2006). Communication à la Commission – La traduction à la Commission. <<http://ec.europa.eu/transparency/regdoc/rep/2/2006/FR/2-2006-1489-FR-F1-1.Pdf>> [24/10/2018].
- Comisión Europea. (2014). Traduction et multilinguisme. Luxembourg: Office des publications de l'Union européenne.
- Čulo, O., Gutermuth, S., Hansen-Schirra, S., & Nitzke, J. (2014). The Influence of Post-Editing on Translation Strategies. In S. O'Brien, L. Winther Balling, M. Carl, M. Simard & L. Specia (Eds.), *Post-Editing of Machine Translation: Processes and Applications* (pp. 200–218). Newcastle upon Tyne: Cambridge Scholars Publishing.
- De Almeida, G. & O'Brien. (2010). Analysing Post-Editing Performance: Correlations with Years of Translation Experience. *Proceedings of the 14th Annual Conference of the European Association for Machine Translation*.

- Dirección General de Traducción. (2015). DGT Translation Quality Guidelines. <http://ec.europa.eu/translation/maltese/guidelines/documents/dgt_translation_quality_guidelines_en.pdf> [24/10/2018].
- European Union External Action. (2016). Blue book traineeship at the EEAS. <https://eeas.europa.eu/headquarters/headquarters-homepage_en/8354/Blue%20book%20traineeship%20at%20the%20EEAS> [24/10/2018].
- Guerberof Arenas, A. (2009). Productivity and quality in the post-editing of outputs from translation memories and machine translation. *The International Journal of Localization*, 7(1), 11-21.
- Koponen, M. (2012). Comparing Human Perceptions of Post-editing Effort with Post-editing Operations. *Proceedings of the Seventh Workshop on Statistical Machine Translation*, 181-190.
- Koponen, M. & Salmi, L. (2015). On the correctness of machine translation: A machine translation post-editing task. *The Journal of Specialised Translation*, 23, 118-138.
- Koskinen, K. (2008). Translating Institutions – An Ethnographic Study of EU Translation. London/New York: Routledge.
- Martikainen, H. & Kübler, N. (2016). Ergonomie cognitive de la post-édition de traduction automatique : enjeux pour la qualité des traductions. *ILCEA-Revue de l'Institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie*, 27, 1-17.
- O'Brien, S. (2011). Towards predicting post-editing productivity. *Springer*, 25, 197-215.
- Peraldi, S. (2016). De la traduction automatique brute à la post-édition professionnelle évoluée : le cas de la traduction financière. *Revue française de linguistique appliquée*, 21, 67-90.
- Santos, F. (2017). Le kappa de Cohen : un outil de mesure de l'accord inter-juges sur des caractères qualitatifs.
- Shuttleworth, M. (2018). Tercera variable. <<https://explorable.com/es/tercera-variable>> [24/10/2018].
- Svoboda, T., Biel, Ł., & Łoboda, K. (2017). Quality aspects in institutional translation. Berlin: Language Science Press.
- Toudic, D., Hernandez Morin, K., Moreau, F., Barbin, F., & Phuez, G. (2014). Du contexte didactique aux pratiques professionnelles: proposition d'une grille multicritères pour l'évaluation de la qualité en traduction spécialisée. *ILCEA-Revue de l'Institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie*, 19, 1-19.